**Building Tesseract OCR (From Source)**

April 6 2021

Tesseract 4.1.1 Leptonica 1.74.4

***# Remove Existing Installation***

1. To remove existing tesseract-ocr following command is used:

sudo apt-get remove tesseract-ocr

2. Following command is used to remove the tesseract-ocr package along with its dependencies:

sudo apt-get remove --auto-remove tesseract-ocr

***# Uninstalling Tesseract that is built from source***

Usually you can just use:

make uninstall

or

sudo make uninstall

***# Install Leptonica***

1. Download latest stable version

<http://leptonica.org/download.html>

Inside Leptonica directory

2. sudo apt-get install libjpeg-dev libpng-dev

3. sudo apt install libpng-dev libtiff5-dev libwebp-dev libopenjp2-7-dev libgif-dev

4. sudo ./configure

5. sudo make

6. sudo make install

***# Install Tesseract dependencies***

$ sudo apt-get update -y

$ sudo apt-get install automake

$ sudo apt-get install -y pkg-config

$ sudo apt-get install -y libsdl-pango-dev

$ sudo apt-get install -y libicu-dev

$ sudo apt-get install -y libcairo2-dev

$ sudo apt-get install bc

1. Find latest stable version for tesseract

<https://github.com/tesseract-ocr/tesseract/releases>

***# Need to run the following commands from the tesseract-4.1.1 directory to install tesseract:***

$ ./autogen.sh

$ ./configure

$ make

$ sudo make install

$ sudo ldconfig

$ make training

$ sudo make training-install

***# Check***

tesseract --version

***# Get Eng Trained data and orientation detection data***

$ wget https://github.com/tesseract-ocr/tessdata/blob/master/eng.traineddata

$ wget https://github.com/tesseract-ocr/tessdata/blob/master/osd.traineddata

export TESSDATA\_PREFIX=usr/local/share/tessdata

To check:

echo $ TESSDATA\_PREFIX

Should give usr/local/share/tessdata

***# Now you can list the languages in your tesseract using the following command:***

$ tesseract --list-langs

You can see the output as following:

List of available languages (2):

eng

osd

Reference:

<https://medium.com/quantrium-tech/installing-tesseract-4-on-ubuntu-18-04-b6fcd0cbd78f>

***#***

Copy eng.traineddata and osd.traineddata to usr/local/share/tessdata

To copy

sudo cp tessdata/\* /usr/local/share/tessdata

To change permissions

chmod +rwx filename ( + → give permissions)

chmod -rwx filename ( - → remove permissions)

Give executable permissions to eng.traineddata and osd.traineddata

***##### Restart machine for changes to reflect ###***